

COPY

"EXPRESS MAIL" Mailing Label Number EL 932952245 US

EXPRESS MAIL LABEL NO. EL692234990US

PATENT APPLICATION
Atty. Docket No.: MDS-022 (6219/6)

A SPECTROSCOPIC SYSTEM EMPLOYING A PLURALITY OF DATA TYPES

Cross-Reference to Related Applications

This application is a continuation-in-part of U.S. patent application Serial No. 09/470,071, filed December 22, 1999, which claims priority to and the benefit of U.S. provisional patent application Serial No. 60/113,761, filed December 23, 1998, and this application is related to the U.S. patent application entitled, "System for Normalizing Spectra" and identified by Attorney Docket Number MDS-020, filed on even date herewith, and to the U.S. patent application entitled, "Spectral Data Classification Of Samples" and identified by Attorney Docket Number MDS-021, filed on even date herewith. All of the above applications are assigned to the common assignee of this application, and are hereby incorporated by reference.

Government Rights

This invention was made with government support under Contract No. CA 66481 awarded by National Cancer Institute, NIH. The government may have certain rights in the invention.

Field of the Invention

This invention relates generally to spectroscopic systems. More particularly, in one embodiment, the invention relates to a spectroscopic system for characterizing test specimens using a plurality of spectral data types.

Background of the Invention

Spectral analysis of biological specimens has been used for disease diagnosis. In general, spectra are recorded as values of amplitude, typically measured as a response to an excitation, as

a function of wavelength (or the inverse of wavelength, namely frequency). In the field of spectral analysis, different kinds of information are conveyed by different spectral types. For example, fluorescence spectra are recorded using a source of excitation illumination that is absorbed by a specimen and that causes the specimen to emit a fluorescence spectrum that depends in part on the transfer of energy within and among atoms and/or molecules in the specimen. An illumination source for use in observing and recording fluorescence spectra generally operates at a selected monochromatic wavelength, or a narrow range of wavelengths. Different sources of illumination that operate at different wavelengths can excite different constituents of a specimen. In addition, different sources of illumination that operate at different wavelengths can excite the same constituent with different efficiencies. Thus, a fluorescence spectrum can depend on both the excitation wavelength that is used to illuminate a specimen and on the composition of the specimen itself. Other effects also play a role in determining a fluorescence spectrum, for example, instrumental effects, effects relating to polarization of the illumination, or thermal effects.

Some progress has been made in using various optical spectral methods for analysis of test specimens, including biological specimens. However, the wide variety of physical and chemical influences present in a test specimen that play roles in determining an observed spectrum make difficult both the choice of a suitable illumination source, and the interpretation of the resulting spectrum. This is in part true because there are so many influences on the kind and amount of information that an optical spectrum conveys that it is hard to find clear cause and effect relationships among the multitude of competing influences.

Summary of the Invention

The invention provides methods of determining the disease state of a biological specimen based upon a reflectance spectrum residual derived by subtracting from a reflectance spectrum obtained from a test specimen an average reflectance spectrum obtained from a plurality of healthy specimens. In a preferred embodiment, members of the plurality of healthy specimens are determined to be healthy based upon the fluorescence spectra emitted by those samples. Typically, the specimen to be tested exhibits a fluorescence spectrum that is not characteristic of healthy tissue.

The reflectance spectrum residual provides a criterion for diagnostic classification of a specimen that is judged to be indeterminate in classification by its fluorescence spectrum alone. Accordingly, methods of the invention resolve diagnostic ambiguities created when a specimen produces a fluorescence spectrum that is not characteristic of a healthy tissue. Similarly, methods of the invention are also useful to resolve diagnostic ambiguities created when the fluorescence spectrum of a test specimen is not characteristic of any known disease state.

In one aspect, the invention provides a method of determining a condition of a test specimen. The method comprises recording both fluorescence and reflectance spectra from specific specimen. The method then comprises identifying a plurality of specimens having a fluorescence spectrum characteristic of a known condition; obtaining an average reflectance spectrum based upon the plurality of first test specimens; obtaining a reflectance spectrum from a test specimen that produces a fluorescence spectrum that is not characteristic of the known condition; and obtaining a reflectance spectrum residual by subtracting the average reflectance spectrum from the reflectance spectrum obtained from the test specimen. Determination of the condition of the test specimen is based upon the reflectance spectrum residual. In a preferred

embodiment, the condition of the test specimen is based upon an amplitude of one or more features of the reflectance spectrum residual. For purposes of the invention a "condition" is a state of disease, including a healthy state or simply the physiological makeup of the specimen and/or the patient from whom it was obtained.

5 In one embodiment, the plurality of specimens producing the average reflectance spectrum comprises tissue specimens and the test specimen is a tissue specimen of the same type. In one embodiment, the tissue specimens are human cervical tissue specimens, the condition of which is healthy, and the condition of the test specimen is determined by methods of the invention. In one embodiment, cervical tissue producing the average reflectance spectrum are selected from normal squamous tissue, metaplasia, mild cervical intraepithelial neoplasia (CIN I), and moderate to severe cervical intraepithelial neoplasia (CIN II/III). In another embodiment, methods of the invention further comprise obtaining additional optical information from the test specimen, and evaluating the additional optical information in comparison to the fluorescence spectrum and the reflectance spectrum from the test specimen to determine the condition of the test specimen. In one embodiment, the additional optical information is video information. In another embodiment, the additional optical information is an optical image.

In one aspect, the invention relates to a spectroscopic system for determining a condition of a test specimen. The system comprises a data collection module that collects a fluorescence spectrum characteristic of a known condition from each of a plurality of first specimens, observes
20 a reflectance spectrum from each member of the plurality, and observes a reflectance spectrum from a test specimen that is observed to produce a second fluorescence spectrum that is not characteristic of the known condition. The system further comprises a computation module that

5 reflectance spectrum residual.

Parameter	Value	Parameter	Value
Mean	1.00	Standard deviation	0.00
Minimum	0.00	Maximum	1.00
Range	1.00	Skewness	0.00
Kurtosis	0.00	Normality test	0.00
Shapiro-Wilk	0.99	Anderson-Darling	0.00
Kolmogorov-Smirnov	0.00	Chi-square	0.00
Likelihood ratio	0.00	Portmanteau	0.00
Q-Q plot	0.00	Normal Q-Q plot	0.00
Sample size	1000	Number of observations	1000

Brief Description of the Drawings

The objects and features of the invention can be better understood with reference to the drawings described below, and the claims. The drawings are not necessarily to scale, emphasis instead generally being placed upon illustrating the principles of the invention. In the drawings, like numerals are used to indicate like parts throughout the various views.

FIG. 1 shows an exemplary spectroscopic system that employs a plurality of spectral types according to an illustrative embodiment of the invention;

FIG. 2 shows an exemplary operational block diagram of the spectroscopic system of FIG. 1;

FIG. 3 is a detailed schematic flow diagram showing exemplary steps of combining a fluorescence spectrum analysis described more fully in FIG. 6 with a reflectance spectrum analysis described more fully in FIG. 7 to perform tissue characterization according to an illustrative embodiment of the invention;

FIG. 4 is a diagram showing an exemplary fluorescence spectrum recorded using the spectroscopic system of FIG. 1;

FIG. 5 is a diagram showing an exemplary reflectance spectrum recorded using the spectroscopic system of FIG. 1;

FIG. 6 is a schematic flow diagram depicting an analysis of a fluorescence spectrum in a system according to an illustrative embodiment of the invention; and

FIG. 7 is a schematic flow diagram depicting an analysis of a reflectance spectrum in a system according to an illustrative embodiment of the invention.

Detailed Description

The invention relates to the use of multiple types of spectral data for determining the status of a test specimen. In the present invention, a first spectral observation is used to distinguish between a first condition and a plurality of other conditions, and a second observation using information obtained from a different type of spectral data is used to distinguish among the plurality of other conditions. The invention will be described in terms of embodiments that relate to the use of multiple optical spectra in characterization systems and methods, particularly in the area of medical diagnostics, and especially as such use relates to the analysis of spectra obtained from human cervical tissue in the detection of cervical cancer. However, the invention has applicability generally in the area of characterization of test specimens based on analysis of a plurality of optical spectra types.

FIG. 1 depicts an exemplary spectroscopic system 100 employing a plurality of spectral data types in methods and systems according to an illustrative embodiment of the invention. The spectroscopic system includes a console 102 connected to a probe 104 by a cable 106. The cable 106 carries electrical and optical signals between the console 102 and the probe 104. The probe 104 accommodates a disposable component 108 which used only once, and discarded after such use. The console 102 and the probe 104 are mechanically connected by an articulating arm 110, which can also support the cable 106. The console 102 contains much of the hardware and the software of the system, and the probe 104 contains the necessary hardware for making suitable spectroscopic observations. The details of the system are further explained in conjunction with FIG. 2.

FIG. 2 shows an exemplary operational block diagram 200 of a spectroscopic system of the type depicted in FIG. 1. According to an illustrative embodiment, the spectroscopic system of FIGs. 1 and 2 is substantially the same as single-beam spectrometer devices, but is adapted to include the features of the invention. The console 102 includes a computer 202 which executes software that controls the operation of the spectroscopic system 100. The software includes one or more modules recorded on machine-readable media, which can be any medium such as magnetic disks, magnetic tape, CD-ROM, semiconductor memory, or the like. Preferably, the machine-readable medium is resident within the computer 202. In alternative embodiments, the machine-readable medium can be connected to the computer 202 by a communication link. In alternative embodiments, one can substitute computer instructions in the form of hardwired logic for software, or one can substitute firmware (i.e., computer instructions recorded on devices such as PROMs, EPROMs or EEPROMs, or the like) for software. The term machine-readable instructions as used herein is intended to encompass software, hardwired logic, firmware and the like.

The computer 202 is a general purpose computer. The computer 202 can be an embedded computer, or a personal computer such as a laptop or desktop computer, that is capable of running the software, issuing suitable control commands, and recording information in real time. The computer 202 has a display 204 for reporting information to an operator of the spectroscopic system 100, a keyboard 206 for enabling the operator to enter information and commands, and a printer 208 for providing a print-out, or permanent record, of measurements made by the spectroscopic system 100 and for printing diagnostic results, for example, for inclusion in the chart of a patient. As described below in more detail, in an illustrative embodiment of the

invention, some commands entered at the keyboard, enable a user to select a particular spectrum for analysis or to reject a spectrum, and to select particular segments of a spectrum for normalization. Other commands enable a user to select the wavelength range for each particular segment and to specify both wavelength contiguous and non-contiguous segments.

5 The console 102 also includes an ultraviolet (UV) source 210 such as a nitrogen laser or a frequency-tripled Nd:YAG laser, a white light source 212 such as one or more Xenon flash lamps, and control electronics 214 for controlling the light sources both as to intensity and as to the time of onset of operation and the duration of operation. One or more power supplies 216 are included in the console 102, to provide regulated power for the operation of all of the components. The console 102 also includes at least one spectrometer and at least one detector (spectrometer and detector 218) suitable for use with each of the light sources. In some embodiments, a single spectrometer can operate with both the UV light source and the white light source. In some embodiments, the same detector can record UV and white light signals, and in some embodiments different detectors are used for each light source.

10 The console 102 also includes coupling optics 220 to couple the UV illumination from the UV light source 210 to one or more optical fibers in the cable 106 for transmission to the probe 104, and for coupling the white light illumination from the white light source 212 to one or more optical fibers in the cable 106 for transmission to the probe 104. The console 102 also includes coupling optics 222 to couple the spectral response of a specimen to UV illumination
20 from the UV light source 210 observed by the probe 104 and carried by one or more optical fibers in the cable 106 for transmission to the spectrometer and detector 218, and for coupling the spectral response of a specimen to the white light illumination from the white light source 212

observed by the probe 104 and carried by one or more optical fibers in the cable 106 for transmission to the spectrometer and detector 218. The console 102 includes a footswitch 224 to enable an operator of the spectroscopic system 100 to signal when it is appropriate to commence a spectral observation by stepping on the switch. In this manner, the operator has his or her hands free to perform other tasks, for example, aligning the probe 104.

The console 102 includes a calibration port 226 for calibrating the optical components of the spectrometer system. The operator places the probe 104 in registry with the calibration port 226 and issues a command that starts the calibration operation. In the calibration operation, a calibrated light source provides illumination of known intensity as a function of wavelength as a calibration signal. The probe 104 detects the calibration signal. The probe 104 transmits the detected signal through the optical fiber in the cable 106, through the coupling optics 222 to the spectrometer and detector 218. A test spectral result is obtained. A calibration of the spectral system is computed as the ratio of the amplitude of the known illumination at a particular wavelength divided by the test spectral result at the same wavelength.

The probe 104 includes probe optics 230 for illuminating a specimen to be analyzed with UV and white light from the UV source 210 and the white light source 212, and for collecting the fluorescent and backscatter (or reflectance) illumination from the specimen that is being analyzed. The probe includes a scanner assembly 232 that provides illumination from the UV source 210 in a raster pattern over a target area of the specimen of cervical tissue to be analyzed. The probe includes a video camera 234 for observing and recording visual images of the specimen under analysis. The probe 104 includes a targeting source 236, which can be used to determine where on the surface of the specimen to be analyzed the probe 104 is pointing. The

probe 104 also includes a white light illuminator 238 to assist the operator in visualizing the specimen to be analyzed. Once the operator aligns the spectroscopic system and depresses the footswitch 224, the computer 202 controls the actions of the light sources 210, 212, the coupling optics 220, the transmission of light signals and electrical signals through the cable 106, the operation of the probe optics 230 and the scanner assembly 232, the retrieval of observed spectra via the cable 106, the coupling of the observed spectra via the coupling optics 222 into the spectrometer and detector 218, the operation of the spectrometer and detector 218, and the subsequent signal processing and analysis of the recorded spectra.

FIG. 3 is a detailed schematic flow diagram 300 showing exemplary steps of combining the fluorescence spectrum analysis described more fully in FIG. 6 below with the reflectance spectrum analysis described more fully in FIG. 7 below to perform tissue characterization according to an illustrative embodiment of the invention. Step 310 indicates that the results presented in FIG. 6 for analysis of fluorescence spectra from a test specimen of unknown condition or unknown state of health are available. At step 320, the computer 202 determines whether the test specimen can be classified as "normal," or "metaplasia," or can not be classified by fluorescence spectroscopy alone. This process is described in detail at step 665 of FIG. 6 below. As indicated in step 325, a decision is taken as to whether the test specimen has a definitive state of health, for example that the specimen is "normal." If the test specimen can be classified, for example as normal, the method ends at step 330.

In the event that a definitive condition or state of health cannot be ascribed to a test specimen, the computer 202 further analyses information available from a reflectance spectrum or from a plurality of reflectance spectra taken from the test specimen. At step 335, the computer

202 provides reflectance spectra processed according to the systems and methods described in connection with FIG. 7 below.

If the specimen cannot be classified, a mean normalization step is performed by computer 202, as indicated at step 340. This mean normalization step is described in detail at step 755 of FIG. 7 below. The mean normalization is carried out using a plurality of reflectance spectra taken from specimens that are known to represent normal squamous tissue. In one embodiment, a single test specimen is examined at multiple locations, each location measuring approximately one millimeter in diameter. If one or more locations of the test specimen provide fluorescence spectra that indicate that those locations can be classified as representing normal squamous tissue, using the methods and systems described in FIG. 6 below, the reflectance spectra recorded from those locations are used to mean normalize the reflectance spectra obtained from locations that are not capable of being classified as "normal" or "metaplasia" solely on the basis of fluorescence spectra.

As indicated in step 350, the computer 202 can carry out an analysis using a metric as described in detail at FIG. 7, step 760 below, for example using the Mahalanobis distance as a metric in N-dimensional space. In one embodiment, the test reflectance spectra are truncated to the wavelength regions 391 nm to 484 nm, and 532 nm to 625 nm. In one embodiment, the classifications CIN I and CIN II/II are the classifications that are possible for a test spectrum that is neither classified as "normal" nor "metaplasia" by fluorescence spectral analysis. As indicated at step 350, the computer 202 classifies the test specimen as having a condition or state of health selected from CIN I and CIN II/III based on the value of the metric computed by the computer 202, provided that the value of the metric does not exceed a pre-determined maximum value.

At step 360, the computer 202 presents the results of the classification of the test specimen, as a condition or state of health corresponding to one of normal, metaplasia, CIN I and CIN II/III.

FIG. 4 is a diagram 400 showing an exemplary fluorescence spectrum recorded using the spectroscopic system of FIG. 1. In FIG. 4, a curve 410 having an amplitude defined in terms of number of counts, for example a signal strength expressed as numbers of photons, as indicated along the vertical axis 420. An amplitude is plotted against a wavelength of light expressed in nm, as indicated along the horizontal axis 430. Such spectra can be recorded using the methods and systems described in Figs. 1-3 with an illumination source that provides optical excitation comprising a substantially monochromatic source in the ultraviolet portion of the optical spectrum. The exemplary fluorescence spectrum depicted in FIG. 4 was recorded from human cervical tissue using ultraviolet illumination.

FIG. 5 is a diagram 500 showing an exemplary reflectance spectrum recorded using the spectroscopic system of FIG. 1. In FIG. 5, a curve 510 having an amplitude defined in terms of number of counts, for example a signal strength expressed as numbers of photons, as indicated along the vertical axis 520. An amplitude is plotted against a wavelength of light expressed in nm, as indicated along the horizontal axis 530. Such spectra can be recorded using the methods and systems described in Figs. 1-3 with an illumination source that provides optical excitation comprising a broadband source in the visible portion of the optical spectrum. The exemplary reflectance spectrum depicted in FIG. 5 was recorded from human cervical tissue using broadband visible light illumination.

FIG. 6 is a schematic flow diagram 600 depicting an analysis of a fluorescence spectrum in a system according to an illustrative embodiment of the invention. Fluorescence spectra such as the exemplary spectrum of FIG. 4 are suitable for the analysis described hereafter. The spectral analysis involves the measurement of spectra from a test specimen whose condition is to be determined, and employs a plurality of reference spectra taken from one or more specimens having known conditions. Those of ordinary skill in the spectroscopic arts will understand that the reference spectra can be obtained before, after, or at the same operation session as the test spectrum or spectra are obtained. For purposes of exposition, the treatment of the reference spectra will be described first. The treatment of a test spectrum will then be described.

In the present application, the term "characteristic N-dimensional value" should be understood to comprehend whichever of a point, a volume, a surface, a probability distribution or the like is used in the subsequent analysis. In mathematical terms, the "characteristic N-dimensional value" can be understood to be an ordered N-tuple of values, each one of which values can be expressed in its own dimensional units. It is well known in the spectroscopic arts to record one or more reference spectra from specimens that are known to have specific conditions. For example, it is known to record reference spectra from specimens having known states of health or known disease conditions. Reference spectra can be manipulated using the same pre-processing methods that are applied to spectra recorded from test specimens, so as to be able to compare spectra having substantially similar processing histories.

The computer 202 can store in a machine-readable memory for later use the information comprising the various characteristic N-dimensional values that are computed, as well as the

corresponding known condition. The computer 202 can also similarly store in a machine-readable memory the various reference reflectance spectra data that were obtained.

Turning to the discussion of the recording and analysis of a test spectrum, as indicated in FIG. 6 at step 610, an operator collects and records a test fluorescence spectrum or a plurality of
5 test fluorescence spectra from a test specimen of unknown state of health or unknown condition.

Step 615 indicates that the computer 202 can smooth the raw data. Various mathematical smoothing techniques are known in the art, such as computing moving averages, or applying filters to remove noise, for example using well-known convolution methods.

Step 620 indicates that the computer 202 can subtract a background reading, for example a reading taken using the systems and methods of FIGs. 1-3 with a standard specimen, or taken without a test specimen, from the reading observed from a test specimen or a reference specimen. In step 625, the computer 202 can average a plurality of spectral data taken from the same specimen, for example, the computer 202 can average N repeated spectral measurements taken from the same location of a single specimen, where N is an integer greater than 1.

In step 630, the computer 202 can correct the wavelength assigned to a particular amplitude of a spectrum, or can correct a plurality of such wavelengths. The computer 202 can perform such a correction by using a pre-recorded calibration file recorded from an optical source having lines at known wavelengths, for example a spectrum recorded from a lamp having the characteristic emission lines of mercury, or another well known emission source. The computer
20 202 uses the known wavelengths of the characteristic emission lines to determine a relationship between wavelength and a data channel number assigned by the computer 202 and an A/D

converter to the characteristic spectra. The computer 202 can determine a wavelength for each A/D channel and thus knows the correct wavelength to assign to each amplitude.

As indicated in step 635, the computer 202 can correct the spectral data for effects ascribable to open air, such as absorption due to atmospheric constituents, for example gases such as water vapor, carbon dioxide, and others that have strong absorption at characteristic wavelengths.

Step 640 indicates that the computer 202 can correct for instrument-induced features by dividing an observed spectrum by an instrument function. The instrument function is obtained by dividing a known, accurate spectrum, such as a NIST-traceable tungsten lamp spectrum, by the observed spectrum recorded for illumination from such a lamp passing through the instrument of FIGs. 1 and 2. Dividing a first spectrum by a second spectrum involves dividing the respective amplitudes of the first spectrum by the corresponding amplitudes for the second spectrum and recording the result of each such division in a file as a function of the respective wavelengths, such as the instrument function.

Step 645 indicates that the computer 202 can truncate a spectrum to limit the data saved to a file as amplitudes corresponding to the reduced set of wavelengths, rather than all the wavelengths that the spectral instrument of FIGs. 1-2 is capable of recording. The truncation is performed to avoid saving data that has no significance, such as data corresponding to amplitudes having zero intensity for wavelengths beyond some predetermined value. In some embodiments, the operator can use the keyboard 206, or another input mechanism, to indicate one or both wavelength limits of the truncated wavelength range to be retained.

As indicated at step 650, the computer 202 can reduce the number of amplitudes, each amplitude corresponding to a selected wavelength, that are used to characterize a spectrum. In some embodiments, the number of amplitudes used to characterize a spectrum is fifty (50). In principle, as few as three amplitudes at three wavelengths may be sufficient to characterize a
5 fluorescence spectrum as representing a test specimen comprising healthy normal squamous cells as distinguished from other tissue types having other states of health.

As indicated at step 655, the computer 202 can normalize the reduced spectral data using a system and method called normalization using non-uniform segmentation. Normalization using non-uniform segment normalization is described in detail in the co-pending patent application entitled "System for Normalizing Spectra," which application is commonly assigned to the assignee of the present application, and which application is incorporated herein by reference in its entirety.

In one embodiment, where the metric to be used in classifying a test specimen is a Mahalanaobis distance, the computer 202 performs a mathematical procedure intended to guarantee that a matrix used in further analytical steps is capable of being inverted. Matrix inversion is a mathematical process well known in the matrix mathematical arts. In one embodiment, the matrix that is computed is called a Friedman matrix. The data used in the calculation of the Friedman matrix is obtained from reference data recorded and stored in machine-readable format. The step of computing the appropriate Friedman matrix can be
20 performed at any time after the reference spectral data is available.

Computation of the Mahalanobis distance requires the inversion of a weighting matrix. One approach (i.e., Linear Discriminant Analysis) utilizes the inverse of the pooled within-

groups covariance while another method (Quadratic Discriminant Analysis) employs the inverse of each within-group covariance. When a "large" number of wavelengths are used (e.g., for large p), either of the aforementioned matrices can be singular and hence cannot be inverted.

Therefore, an alternate weighting matrix called the Friedman matrix is used.

γ

5 The Friedman matrix is the weighted linear combination of a within-group covariance matrix C_j and the pooled within-group covariance C that is a function of the *Friedman parameters* γ and λ . These parameters are selected from the unit interval $[0,1]$. It is important to note that the Friedman parameters are *not physical quantities whose values carry classification information*. Rather, they are used to insure that the weighting (Friedman) matrix is non-singular. Equation 1 details the p -by- p *Friedman matrix*, which is denoted by Ω .

$$\Omega_j(\gamma, \lambda) = (1 - \gamma) \cdot [(1 - \lambda) \cdot C_j + \lambda \cdot C] + \frac{\gamma}{p} \cdot \text{tr}[(1 - \lambda) \cdot C_j + \lambda \cdot C] \cdot I_{pxp} \quad \text{Eqn. 1}$$

Note that $\text{tr}(\mathbf{A})$ is the trace of the matrix \mathbf{A} , namely the sum of the diagonal elements. Also, I_{pxp} is the p -dimensional identity matrix.

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \Rightarrow \text{tr}(\mathbf{A}) = \sum_{k=1}^n a_{k,k} \quad \text{Eqn. 2}$$

15

$$I_{pxp} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{pxp} \quad \text{Eqn. 3}$$

Let $s(\Lambda_p)$ be a p -dimensional spectrum. The *Mahalanobis distance* d from $s(\Lambda_p)$ to the group mean μ_j with respect to the associated Friedman matrix $\Omega_j(\gamma, \lambda)$ and Friedman parameters γ and λ is given in equation 4 below.

$$d(s(\Lambda_p), \mu_j) = \sqrt{\left| (s(\Lambda_p) - \mu_j) \bullet \Omega_j(\gamma, \lambda)^{-1} \bullet (s(\Lambda_p) - \mu_j)^T \right|} \quad \text{Eqn. 4}$$

Where other metrics are used to classify a test spectrum and a test specimen, the step 670 can be replaced by the appropriate computation.

As indicated at step 660, the computer 202 computes a metric from each test spectrum (or from an average value representing a plurality of spectra, such as performed at step 625) to each characteristic N-dimensional value for the reference spectra recorded. In one embodiment, the metric can be a Mahalanobis distance, as indicated above. In other embodiments, the metric can be the square root of the sum of the squares of the differences in coordinates in the N-dimensional space, or the metric can be the Bhattacharyya distance.

At step 665, the computer 202 examines the results obtained in step 660, and determines a classification for the test specimen, based on the metrics computed with respect to the test fluorescence spectrum and a classification rule or relation. In one embodiment, the classification relation is that the condition of the test specimen is assigned as the condition corresponding to the reference spectrum constellation or set having the Mahalanobis shortest distance, provided that the shortest Mahalanobis distance is less than a pre-determined minimum distance. In one

embodiment, if no Mahalanobis distance is less than a pre-determined minimum distance, the test spectrum is discarded as being incapable of being classified, or indeterminate.

In one embodiment, the test fluorescence spectrum can be compared to multiple sets of reference spectra in a single comparison. In one embodiment, the computations can be repeated
5 for different sets of reference conditions. For example, a test fluorescence spectrum from a test specimen can be compared in a first computation to reference fluorescence spectral data for normal squamous tissue and CIN II/III tissue, compared in a second computation to reference fluorescence spectral data for normal squamous tissue and CIN I tissue, and compared in a third computation to reference fluorescence spectral data for normal squamous tissue and metaplasia tissue. If the test fluorescence spectrum is classified in each of the three comparisons as being more closely related to the reference spectra for normal squamous tissue, the test specimen can be classified by the computer 202 as having normal health, and the classification process is complete. However, if the test fluorescence spectrum is classified in either of the comparisons involving CIN I or CIN II/III as being less closely related to the reference spectra for normal squamous tissue than one of the other conditions or states of health (namely CIN I, or CIN II/III), the computer 202 can report that the state of health of the test specimen is not clearly that of normal squamous tissue, and that further analysis is in order. In one embodiment, the further analysis involves the recording and examination of reflectance spectra.

FIG. 7 is a schematic flow diagram 700 depicting an analysis of a reflectance spectrum in
20 a system according to an illustrative embodiment of the invention. Reflectance spectra such as the exemplary spectrum of FIG. 5 are suitable for the analysis described hereafter. The spectral analysis involves the measurement of spectra from a test specimen whose condition is to be

determined, and employs a plurality of reference spectra taken from one or more specimens having known conditions. Those of ordinary skill in the spectroscopic arts will understand that the reference spectra can be obtained before, after, or at the same operation session as the test spectrum or spectra are obtained.

5 The computer 202 can store in a machine-readable memory for later use the information comprising the various characteristic N-dimensional values that are computed, as well as the corresponding known condition. The computer 202 can also similarly store in a machine-readable memory the various reference reflectance spectra data that were obtained.

Turning to the discussion of the recording an analysis of a test spectrum, as indicated in FIG. 7 at step 710, an operator collects and records a test reflectance spectrum or a plurality of test reflectance spectra from a test specimen of unknown state of health or unknown condition.

Step 715 indicates that the computer 202 can smooth the raw data. Various mathematical smoothing techniques are known in the art, such as computing moving averages, or applying filters to remove noise, for example using well-known convolution methods.

Step 720 indicates that the computer 202 can subtract a background reading, for example a reading taken using the systems and methods of FIGs. 1-3 with a standard specimen, or taken without a test specimen, from the reading observed from a test specimen or a reference specimen. In step 725, the computer 202 can average a plurality of spectral data taken from the same specimen, for example, the computer 202 can average N repeated spectral measurements taken
20 from the same location of a single specimen, where N is an integer greater than 1.

In step 730, the computer 202 can correct the wavelength assigned to a particular amplitude of a spectrum, or can correct a plurality of such wavelengths. The computer 202 can

perform such a correction by using a pre-recorded calibration file recorded from an optical source having lines at known wavelengths, for example a spectrum recorded from a lamp having the characteristic emission lines of mercury, or another well known emission source. The computer 202 uses the known wavelengths of the characteristic emission lines to determine a relationship between wavelength and a data channel number assigned by the computer 202 and an A/D converter to the characteristic spectra. The computer 202 can determine a wavelength for each A/D channel and thus knows the correct wavelength to assign to each amplitude.

As indicated in step 735, the computer 202 can correct the spectral data for effects ascribable to open air, such as absorption due to atmospheric constituents, for example gases such as water vapor, carbon dioxide, and others that have strong absorption at characteristic wavelengths.

Step 740 indicates that the computer 202 can correct for instrument-induced features by dividing an observed spectrum by an instrument function. The instrument function is obtained by dividing a known, accurate spectrum, such as a NIST-traceable tungsten lamp spectrum, by the observed spectrum recorded for illumination from such a lamp passing through the instrument of FIGs. 1 and 2. Dividing a first spectrum by a second spectrum involves dividing the respective amplitudes of the first spectrum by the corresponding amplitudes for the second spectrum and recording the result of each such division in a file as a function of the respective wavelengths, such as the instrument function.

Step 745 indicates that the computer 202 can truncate a spectrum to limit the data saved to a file as amplitudes corresponding to the reduced set of wavelengths, rather than all the wavelengths that the spectral instrument of FIGs. 1-2 is capable of recording. The truncation is

performed to avoid saving data that has no significance, such as data corresponding to amplitudes having zero intensity for wavelengths beyond some predetermined value. In some embodiments, the operator can use the keyboard 206, or another input mechanism, to indicate one or both wavelength limits of the truncated wavelength range to be retained.

5 As indicated at step 750, the computer 202 can reduce the number of amplitudes, each amplitude corresponding to a selected wavelength, that are used to characterize a spectrum. In some embodiments, the number of amplitudes used to characterize a spectrum is fifty (50). In principle, as few as three amplitudes at three wavelengths may be sufficient to characterize a reflectance spectrum as representing a test specimen comprising healthy normal squamous cells as distinguished from other tissue types having other states of health.

10 As indicated at step 755, the computer 202 can normalize the reduced spectral data using a system and method called mean normalization. Mean normalization involves determining a mean value at each wavelength of interest for reference reflectance spectra and subtracting the mean values so determined at a particular wavelength of interest from the amplitude of the test spectrum (or spectra) at the same wavelength.

15 In one embodiment, at step 760, where the metric to be used in classifying a test specimen is a Mahalanaobis distance, the computer 202 performs a mathematical procedure intended to guarantee that a matrix used in further analytical steps is capable of being inverted. Matrix inversion is a mathematical process well known in the matrix mathematical arts. In one
20 embodiment, the matrix that is computed is called a Friedman matrix. The data used in the calculation of the Friedman matrix is obtained from the stored data recorded in machine-readable

format. The step of computing the appropriate Friedman matrix can be performed at any time after the reference spectral data is available.

Computation of the Mahalanobis distance requires the inversion of a weighting matrix.

One approach (i.e., Linear Discriminant Analysis) utilizes the inverse of the pooled within-

5 groups covariance while another method (Quadratic Discriminant Analysis) employs the inverse of each within-group covariance. When a "large" number of wavelengths are used (e.g., for large p), either of the aforementioned matrices can be singular and hence cannot be inverted.

Therefore, an alternate weighting matrix called the Friedman matrix is used.

The Friedman matrix is the weighted linear combination of a within-group covariance matrix C_j and the pooled within-group covariance C that is a function of the *Friedman parameters* γ and λ . These parameters are selected from the unit interval $[0,1]$. It is important to note that the Friedman parameters are *not physical quantities whose values carry classification information*. Rather, they are used to insure that the weighting (Friedman) matrix is non-singular. Equation 5 details the p -by- p *Friedman matrix*, which is denoted by Ω .

15
$$\Omega_j(\gamma, \lambda) = (1 - \gamma) \cdot [(1 - \lambda) \cdot C_j + \lambda \cdot C] + \frac{\gamma}{p} \cdot \text{tr}[(1 - \lambda) \cdot C_j + \lambda \cdot C] \cdot I_{p \times p} \quad \text{Eqn. 5}$$

Note that $\text{tr}(\mathbf{A})$ is the trace of the matrix \mathbf{A} , namely the sum of the diagonal elements. Also, $I_{p \times p}$ is the p -dimensional identity matrix.

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \Rightarrow \text{tr}(\mathbf{A}) = \sum_{k=1}^n a_{k,k} \quad \text{Eqn. 6}$$

$$I_{p \times p} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}_{p \times p} \quad \text{Eqn. 7}$$

Let $s(\Lambda_p)$ be a p -dimensional spectrum. The *Mahalanobis distance* d from $s(\Lambda_p)$ to the group mean μ_j with respect to the associated Friedman matrix $\Omega_j(\gamma, \lambda)$ and Friedman parameters γ and λ is given in equation 8 below.

$$d(s(\Lambda_p), \mu_j) = \sqrt{\left| (s(\Lambda_p) - \mu_j) \bullet \Omega_j(\gamma, \lambda)^{-1} \bullet (s(\Lambda_p) - \mu_j)^T \right|} \quad \text{Eqn. 8}$$

Where other metrics are used to classify a test spectrum and a test specimen, the step 770 can be replaced by the appropriate computation.

As indicated at step 760, the computer 202 computes a metric from each test spectrum (or from an average value representing a plurality of spectra, such as performed at step 725) to each characteristic N-dimensional value for the reference spectra recorded. In one embodiment, the metric can be a Mahalanobis distance, as indicated above. In other embodiments, the metric can be the square root of the sum of the squares of the differences in coordinates in the N-dimensional space, or the metric can be the Bhattacharyya distance.

At step 765, the computer 202 examines the results obtained in step 760, and determines a classification for the test specimen, based on the metrics computed with respect to the test reflectance spectrum and a classification rule or relation. In one embodiment, the classification relation is that the condition of the test specimen is assigned as the condition corresponding to

the reference spectrum constellation or set having the Mahalanobis shortest distance, provided that the shortest Mahalanobis distance is less than a pre-determined minimum distance. In one embodiment, if no Mahalanobis distance is less than a pre-determined minimum distance, the test spectrum is discarded as being incapable of being classified, or indeterminate.

5 In one embodiment, the test reflectance spectrum can be compared to multiple sets of reference spectra in a single comparison. In one embodiment, the computations can be repeated for different sets of reference conditions. For example, a test reflectance spectrum from a test specimen can be compared in a first computation to reference reflectance spectral data for normal squamous tissue and CIN II/III tissue, compared in a second computation to reference reflectance spectral data for normal squamous tissue and CIN I tissue, and compared in a third computation to reference reflectance spectral data for normal squamous tissue and metaplasia tissue. If the test reflectance spectrum is classified in each of the three comparisons as being more closely related to the reference spectra for normal squamous tissue, the test specimen can be classified by the computer 202 as having normal health, and the classification process is complete. However, if the test reflectance spectrum is classified in either of the comparisons involving CIN I or CIN II/III as being less closely related to the reference spectra for normal squamous tissue than one of the other conditions or states of health (namely CIN I, or CIN II/III), the computer 202 can report that the state of health of the test specimen is not clearly that of normal squamous tissue, and that further analysis is in order. In one embodiment, the further analysis involves the recording and
20 examination of reflectance spectra.

Equivalents

While the invention has been particularly shown and described with reference to specific preferred embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

What is claimed is:

10074933-020300